

A QSAR to Predict Plasma Protein Binding of Structurally Diverse Pharmaceutical Compounds

11/18/2009

Christopher Ruark

Assistant Biochemist

Henry M. Jackson Foundation

Air Force Research Laboratory

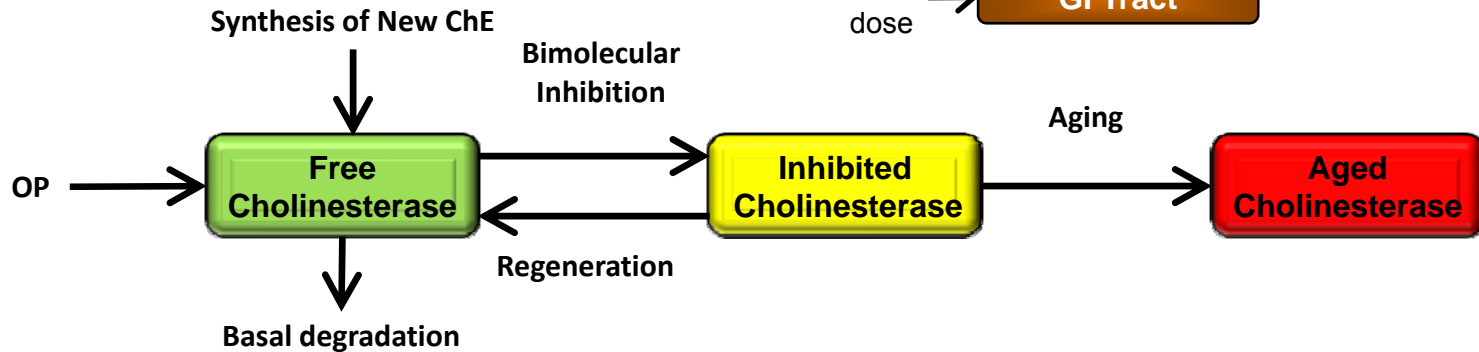
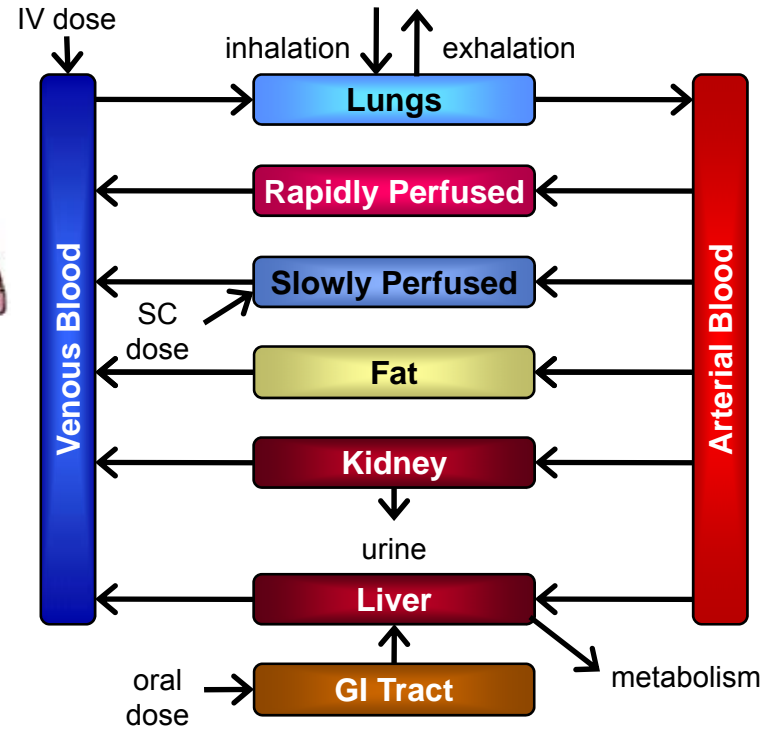
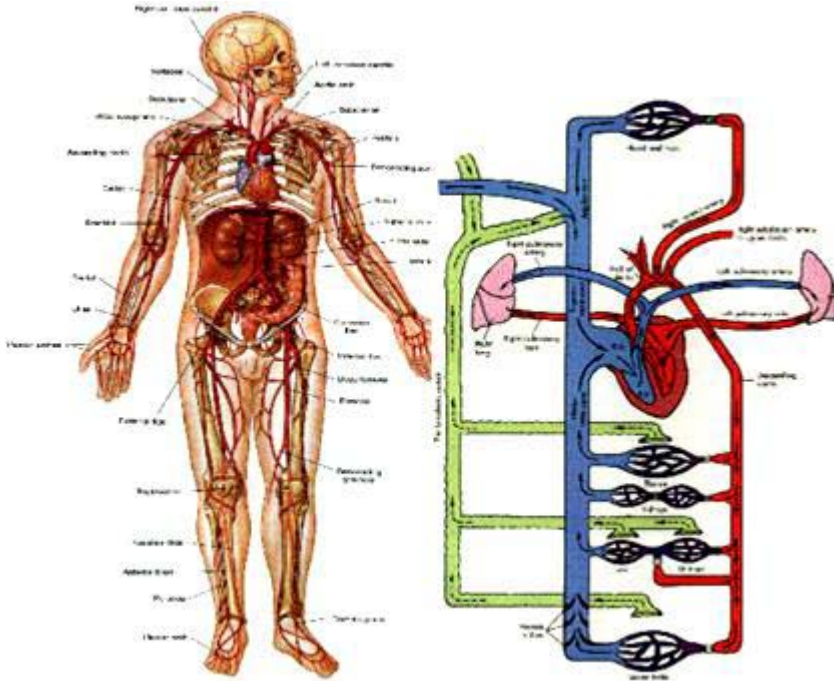
2nd Year Masters Student

Wright State University





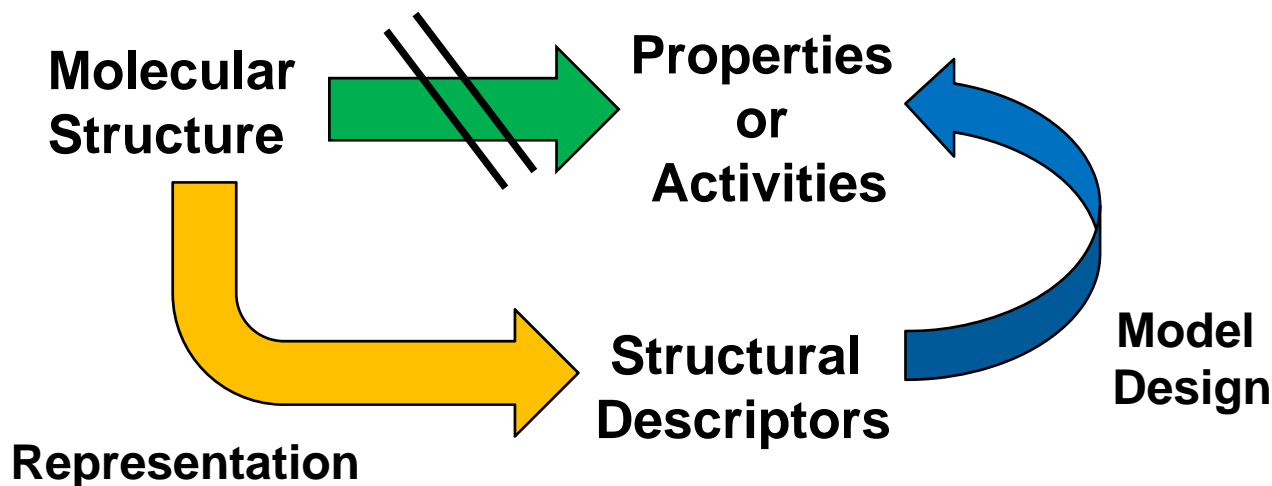
PBPK Modeling





QSAR

A technique used to quantify differences between biological activity and that of a molecular structure

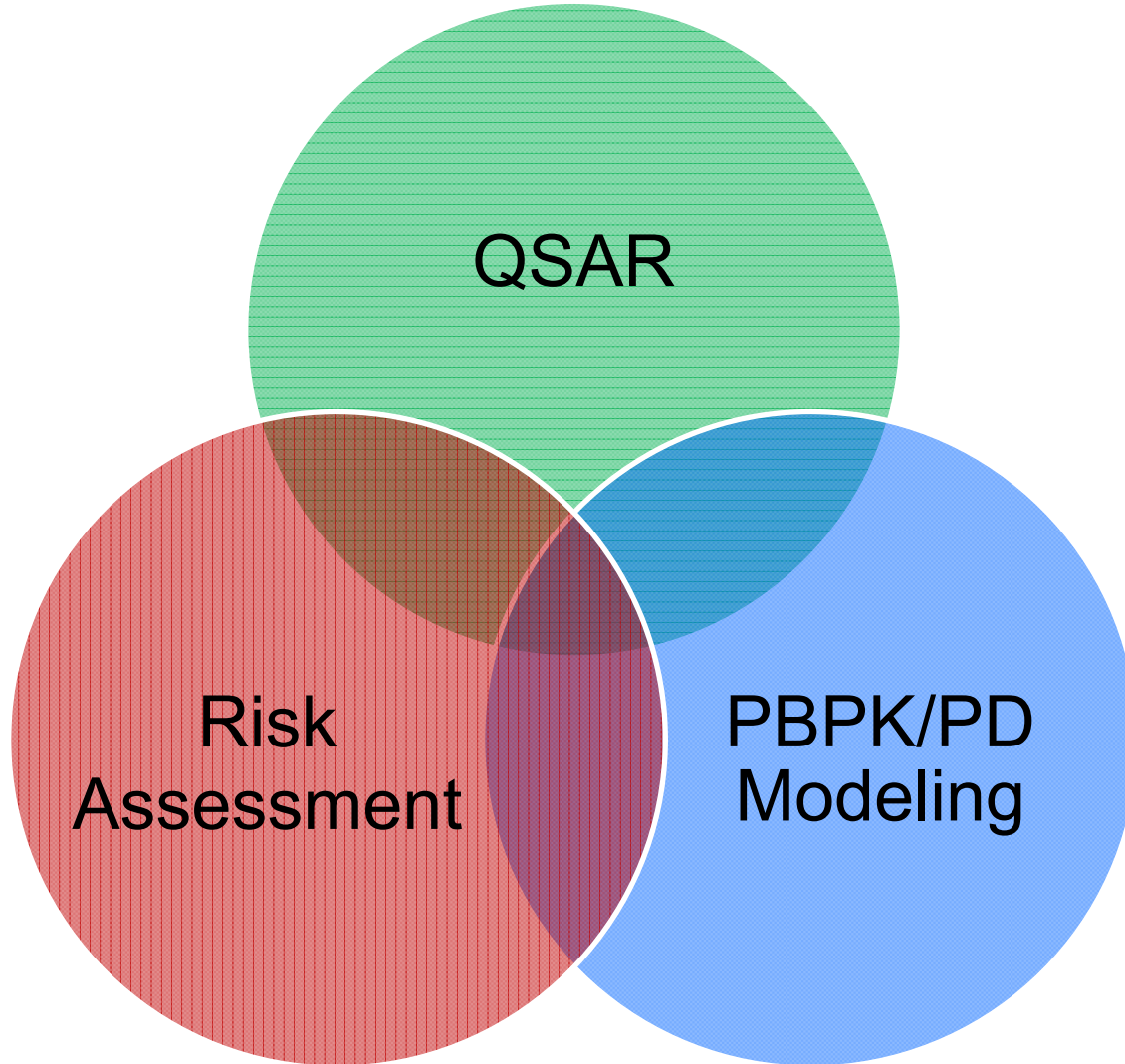


$$\text{Log (PPB)} = F_1x_1 + F_2x_2 + F_3x_3 + F_4x_4 + \dots F_nx_n + \Delta X$$

Fitted by multiple linear regression (MLR)



In Silico Modeling Toolkit

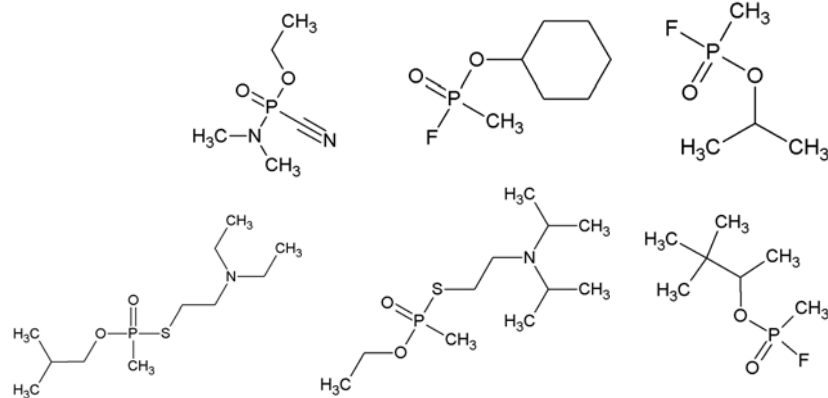




QSAR Endpoints

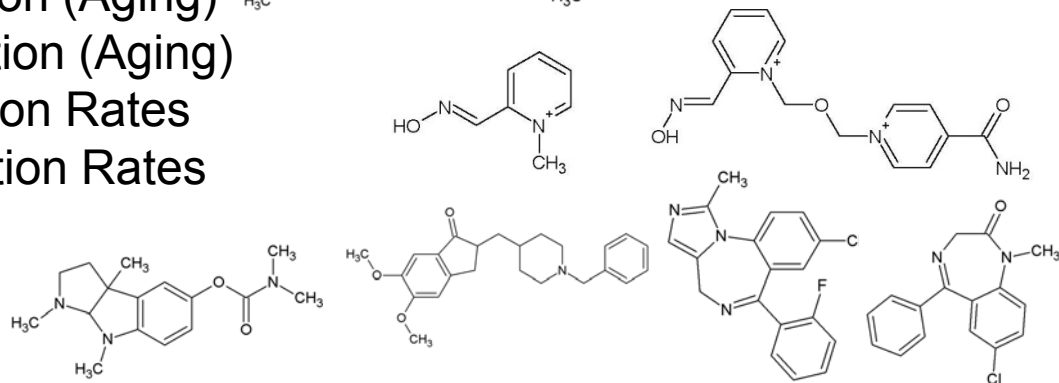
Organophosphates

- Acetylcholinesterase Bimolecular Rates
- Butyrylcholinesterase Bimolecular Rates
- Carboxylesterase Bimolecular Rates
- Trypsin Bimolecular Rates
- Chymotrypsin Bimolecular Rates
- Acetylcholinesterase Dealkylation (Aging)
- Butyrylcholinesterase Dealkylation (Aging)
- Acetylcholinesterase Reactivation Rates
- Butyrylcholinesterase Reactivation Rates



Oximes

- Reactivation Rates
- Dissociation Rates



Reversible Cholinesterase Inhibitors

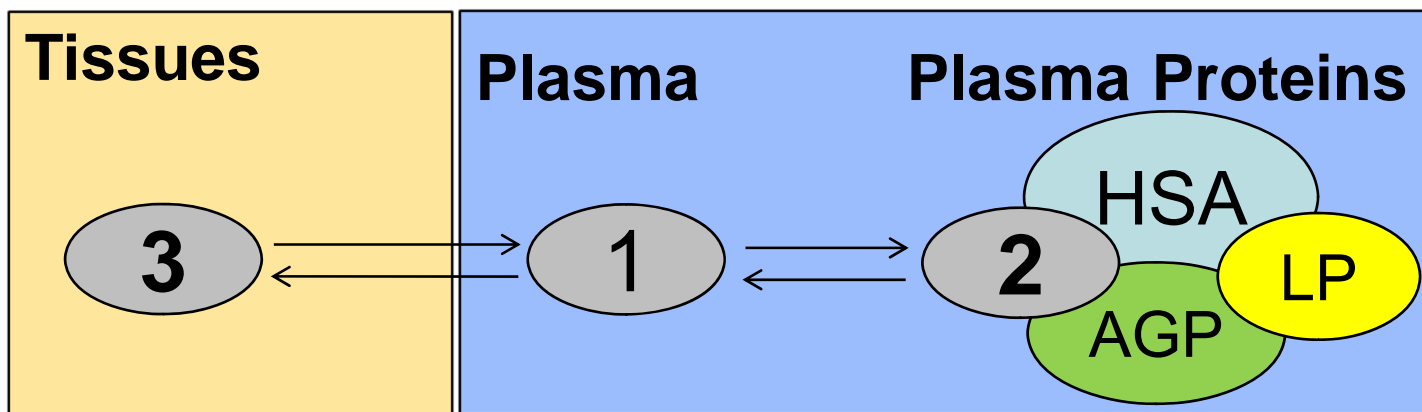
Tissue/Blood Partition Coefficients

Plasma Protein Binding

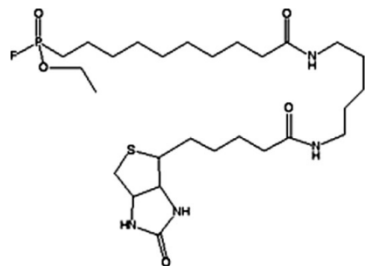
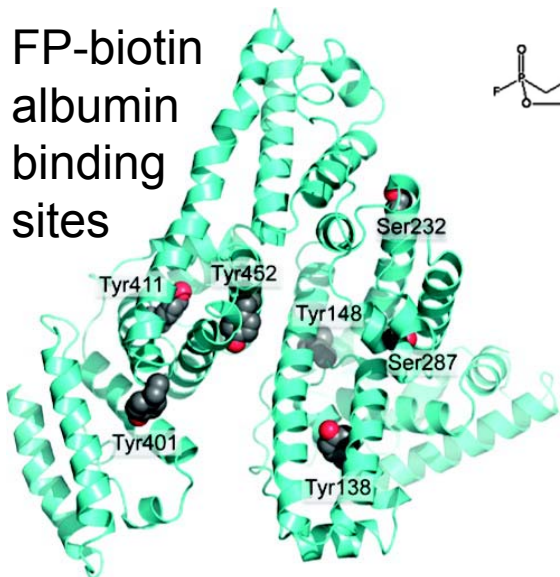
Metabolism



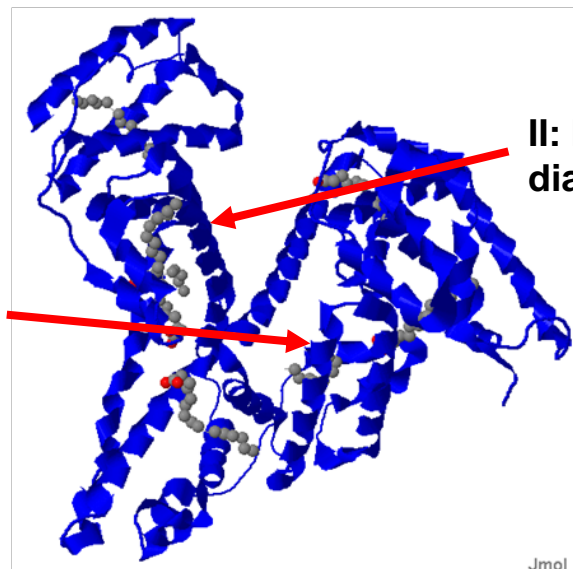
Introduction to Plasma Protein Binding



FP-biotin albumin binding sites



I:
Warfarin site

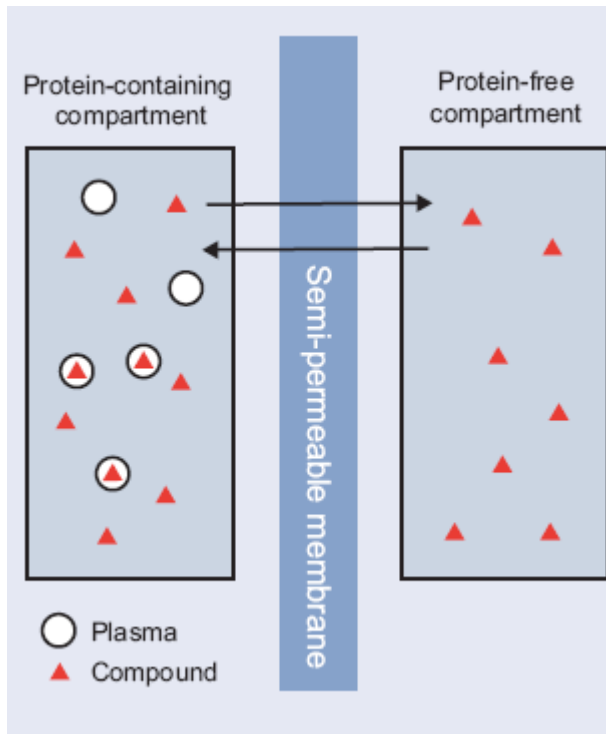


II: Indole-benzo diazepine site

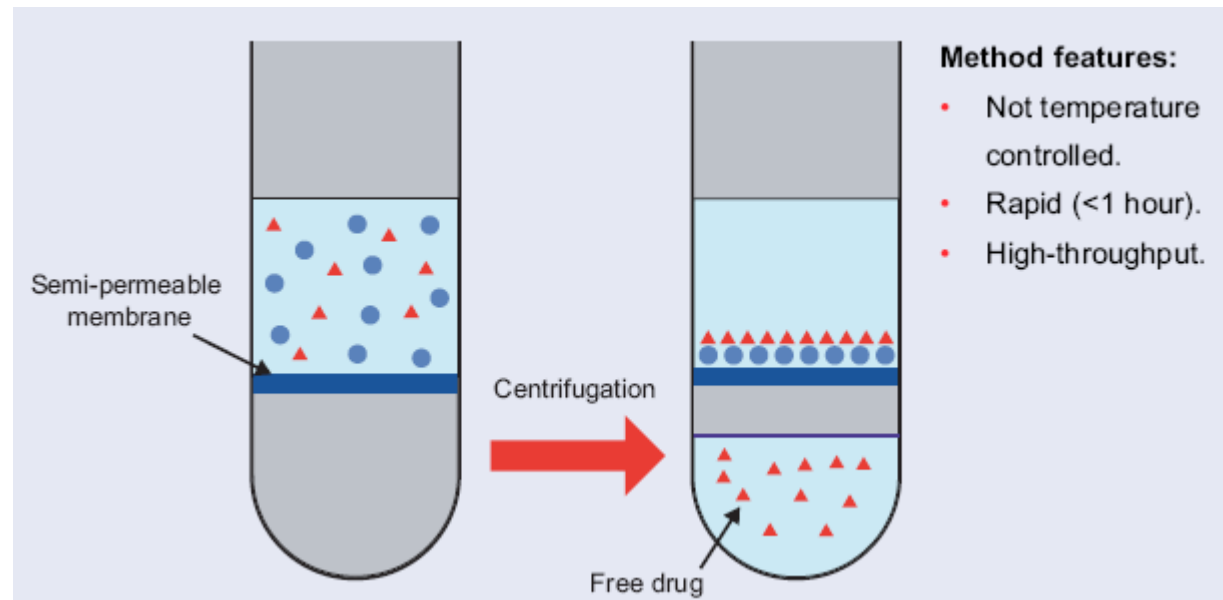
Jmol



Measuring Plasma Protein Binding



Equilibrium dialysis



Ultracentrifugation

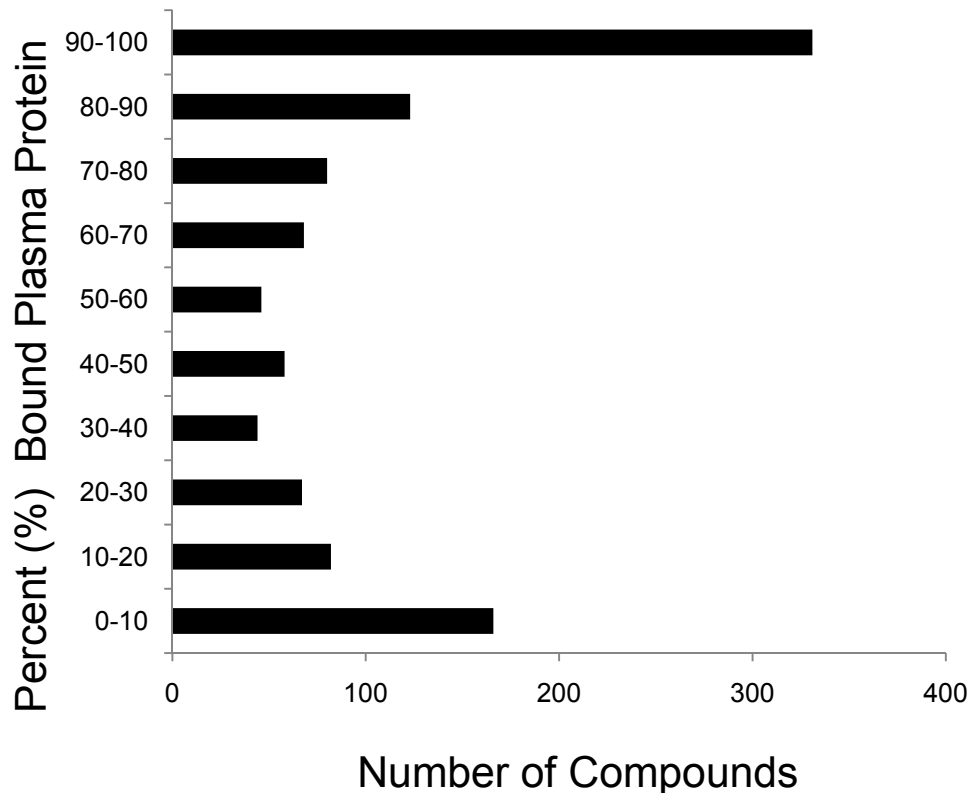


Set of Investigated Compounds

Data on the percent fraction of compounds bound to human plasma proteins was taken from Votano et al. 2006.

Largest published data set available.
High level of chemical diversity.

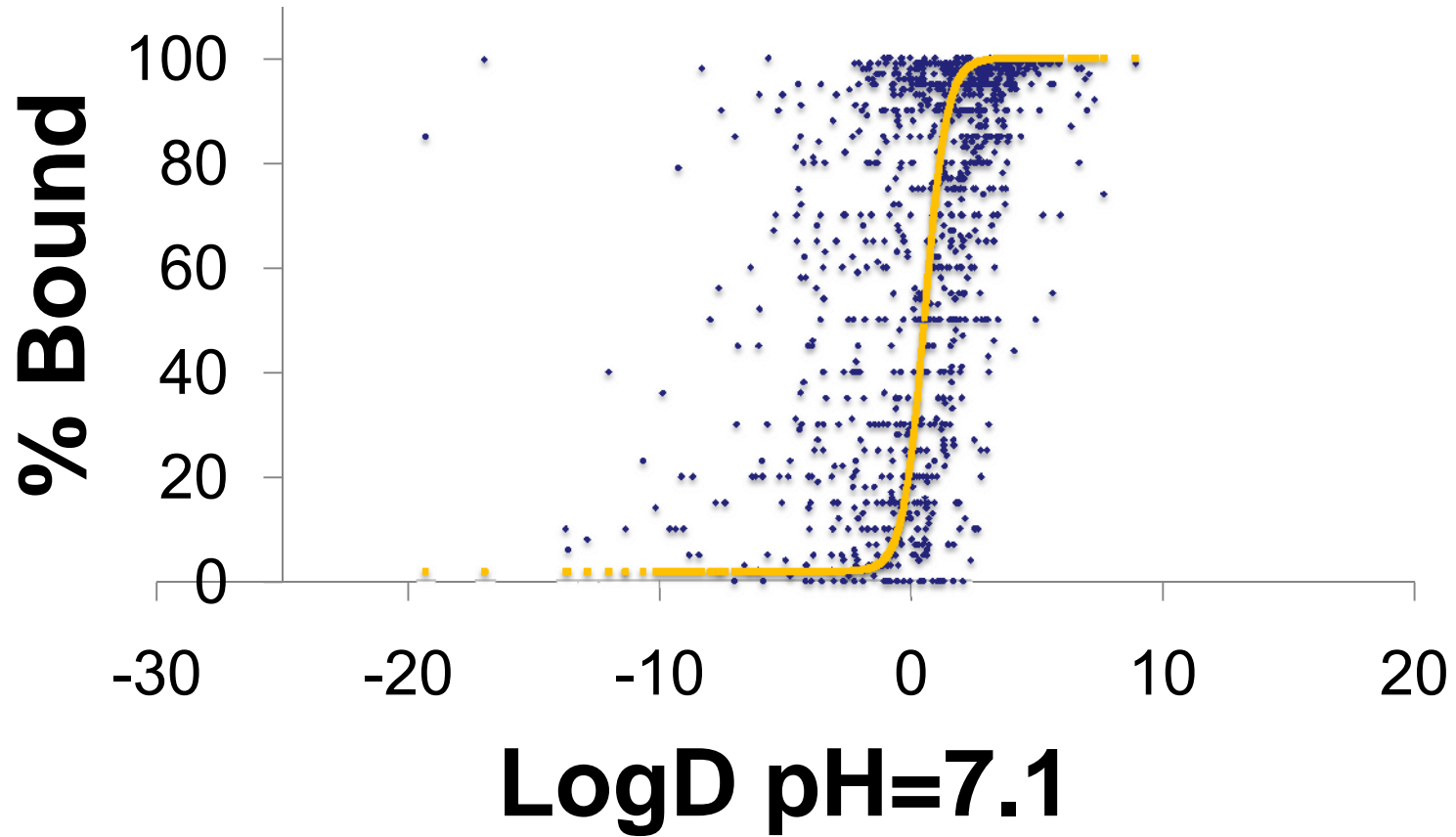
Description	% of database
One or more ring structure	97
One amine	61
Hydroxyl group	39
Amide group	35
Carboxylic acid group	25



Percentage data were converted into an equivalent binding affinity k_A , with the following formula derived from the law of mass: $\log k_A = \log ([f_b] / (1 - [f_b]) - \log[HSA])$



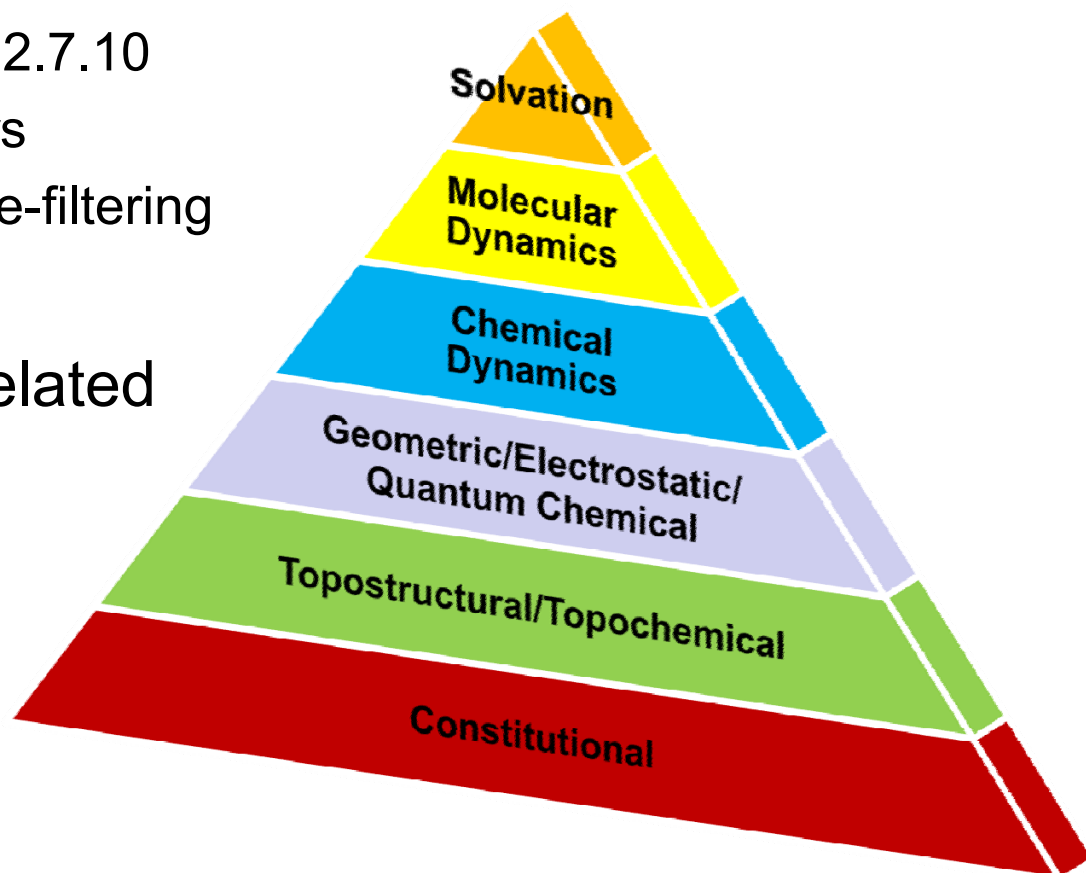
Results





Physiochemical Descriptors

- Ampac 8.0 and Codessa 2.7.10
 - Set of 810 Descriptors
 - 190 removed after pre-filtering
- Highest linearly correlated descriptor
 - Relative number of Carbon atoms

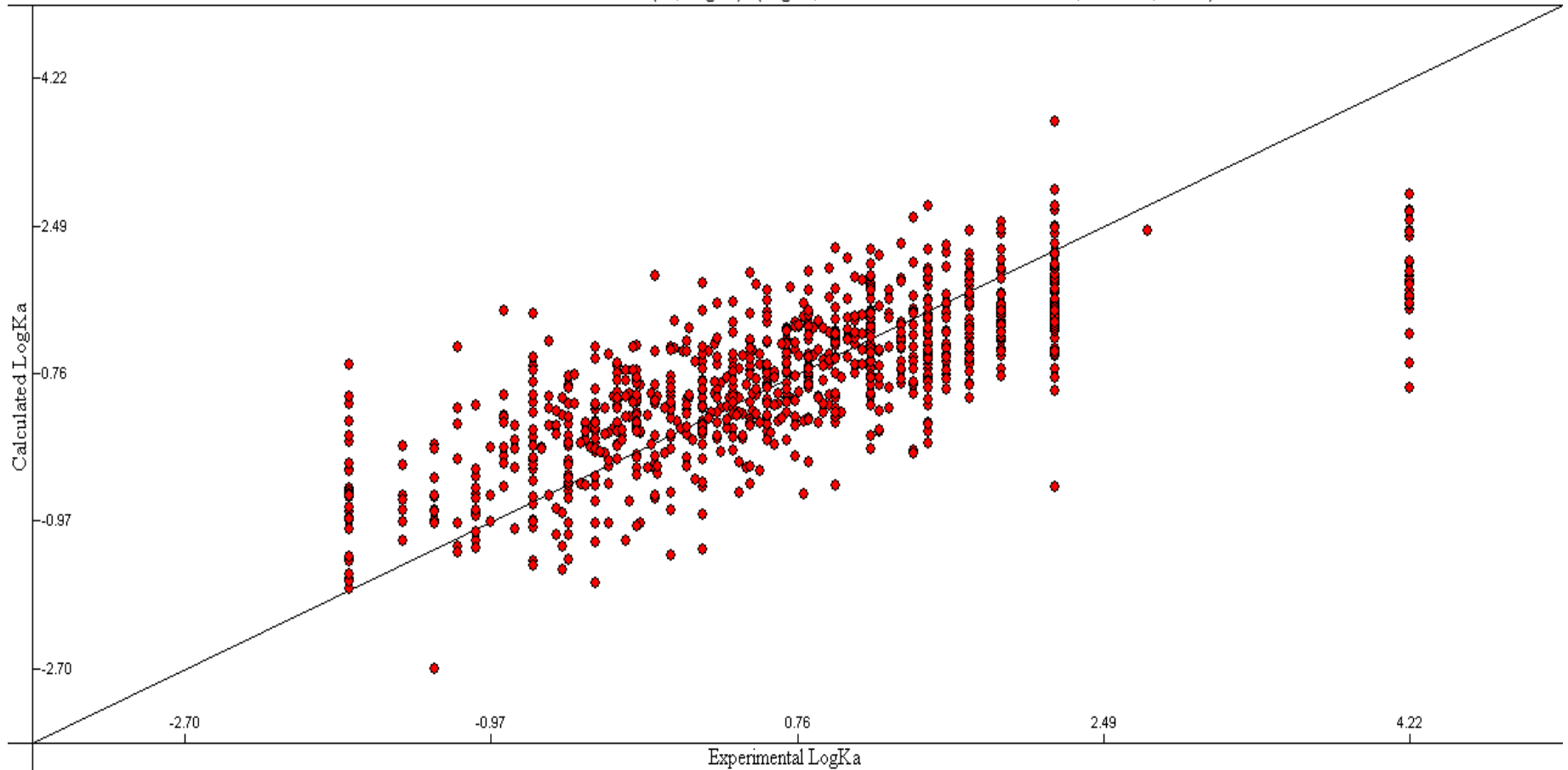




Results : Multi-linear Regression

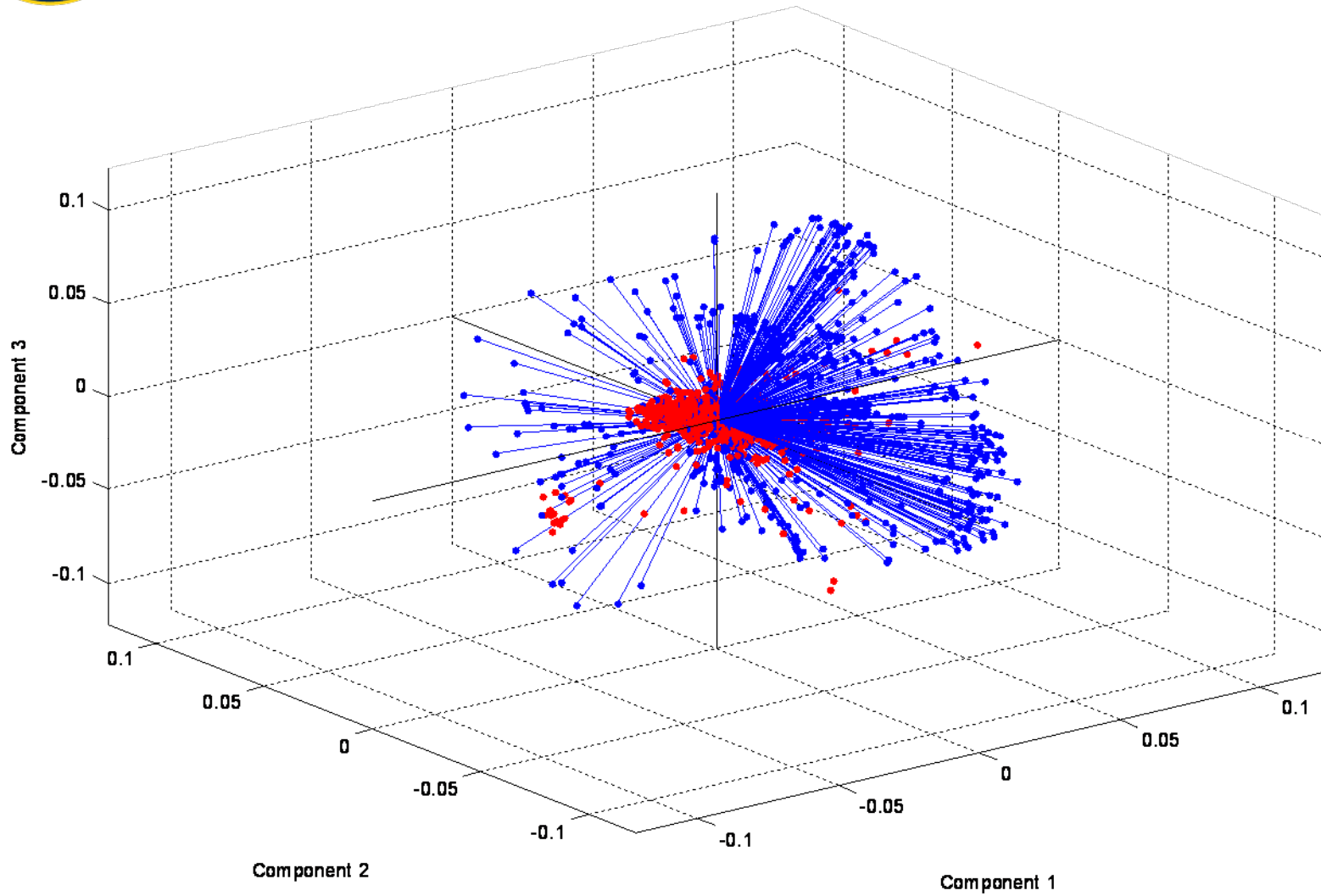
Prediction: TSS 1 (984 structures, 984 plotted, RMSE=0.7711, R2=0.5627)

Corr. Used: "R2=0.5627 F= 30.3406 (40, LogKa)" (LogKa, R2=0.5627 F=30.34 s2=0.6204, 40 descs, TSS 1)



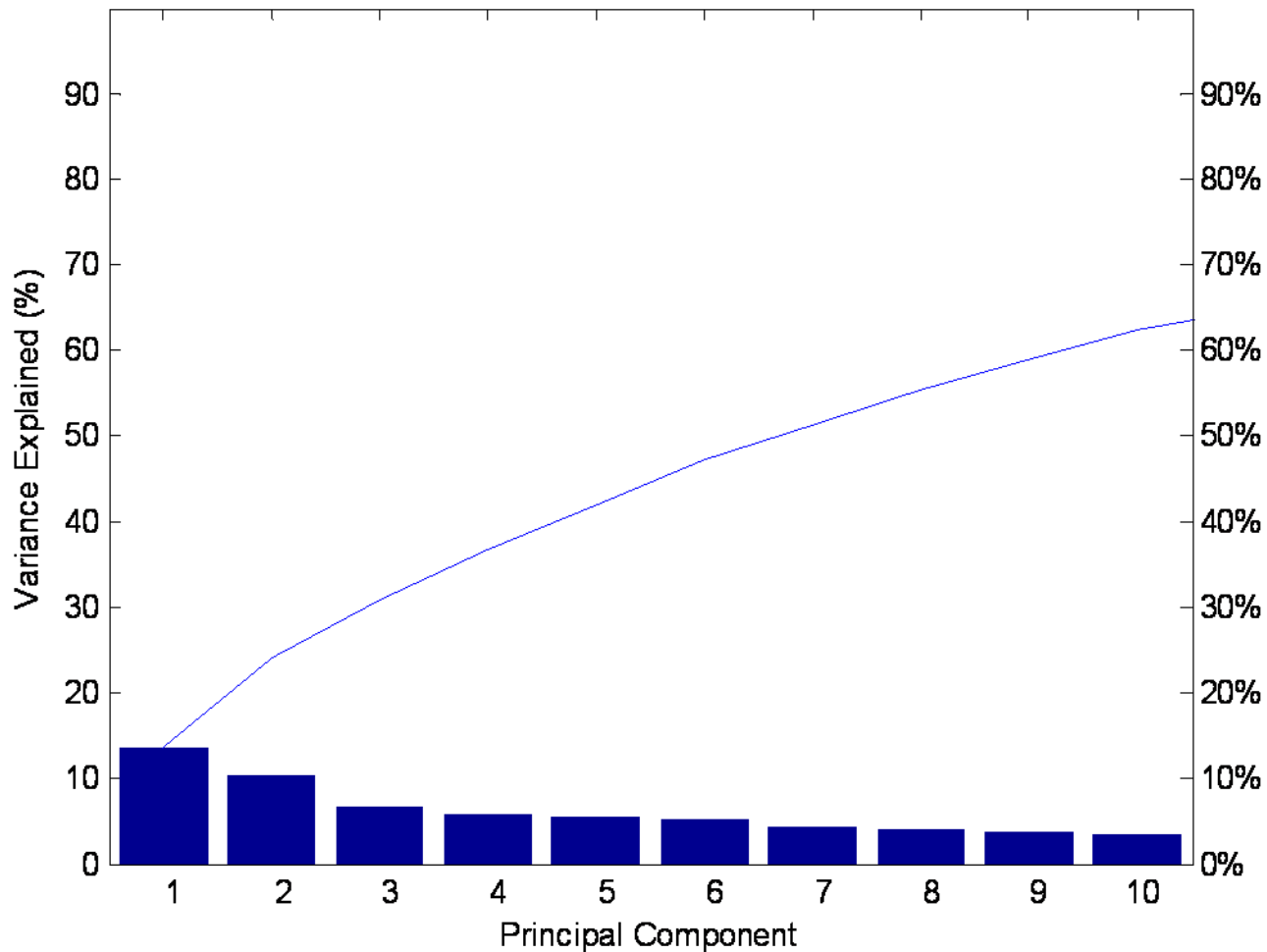


Results: Principle Component 3D Biplot



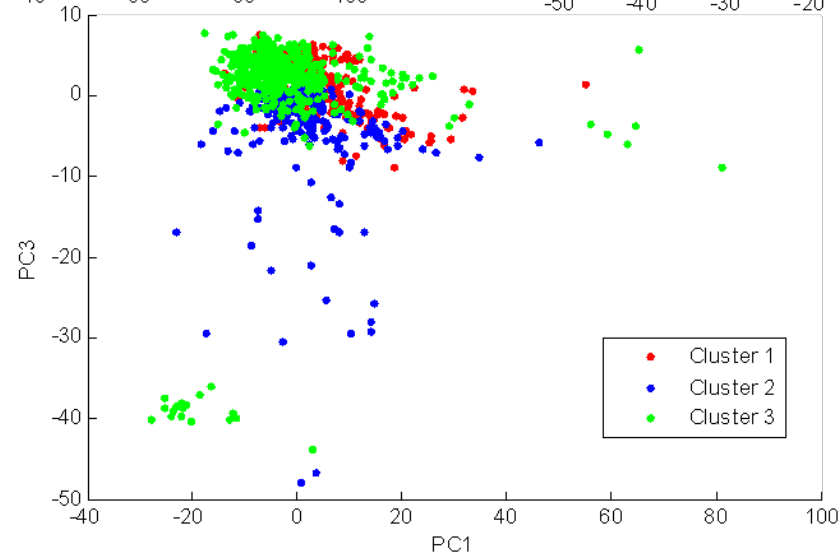
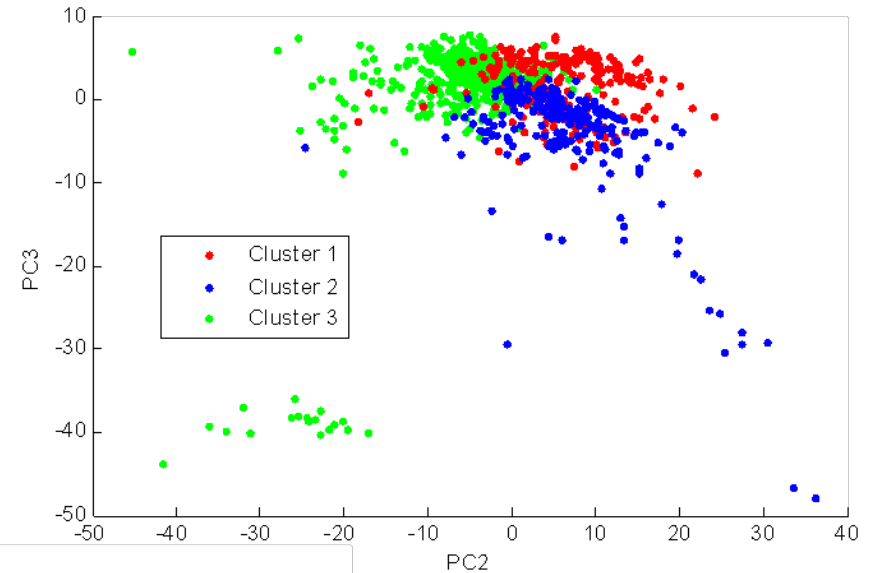
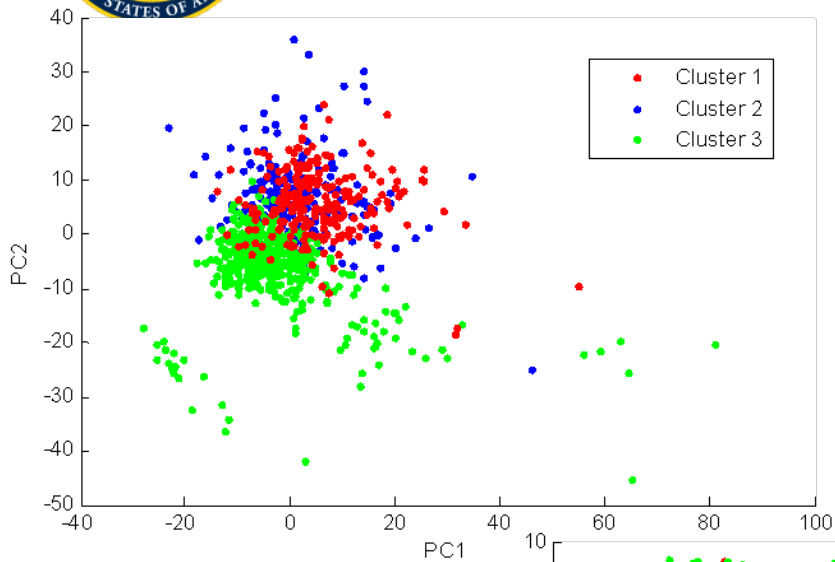


Results : Variance Explained by PC's





Results: PC1 vs. PC2 vs. PC3





Future Directions

- Further analyze data
- Develop randomized test set and AD to evaluate model robustness.
- Introduce new regression techniques to improve correlation.
 - Ridge Regression
 - Neural Networking
 - Support Vector Machine Learning
- Apply predictions to PBPK/PD models



Acknowledgements

Dr. Jeffery Gearhart
Dr. Peter Robinson
Eric Hack
Tammie Covington
Elaine Merrill

Teri Sterner
Emily Fleming
Dr. Kyung Yu
Meghan Makley
Lee Poepelman
Deirdre Mahle
Tim Hansell

This project (2.G 806 AHB C) received support from the Defense Threat Reduction Agency – Joint Science and Technology Office, Basic and Supporting Sciences Division.